Generation of MRS Abstract Predicates from Paninian USR

Sukhada Sukhada
IIT (BHU) Varanasi

Sirisipalli Veera Hymavathi
IIT (BHU), Varanasi

Soma Paul
IIIT Hyderabad

Proceedings of the 30th International Conference on
Head-Driven Phrase Structure Grammar

University of Massachusetts Amherst

Stefan Müller, Elodie Winckel (Editors)

2023

Frankfurt/Main: University Library

pages 122–142

Keywords: Universal Semantic Representation, Minimal Recursion Semantics, Abstract Predicates, Transfer Grammar, Indian Grammatical Tradition, Panini, vivakṣā, and kāraka.

Sukhada, Sukhada, Sirisipalli Veera Hymavathi & Soma Paul. 2023. Generation of MRS Abstract Predicates from Paninian USR. In Stefan Müller & Elodie Winckel (eds.), *Proceedings of the 30th International Conference on Head-Driven Phrase Structure Grammar, University of Massachusetts Amherst*, 122–142. Frankfurt/Main: University Library. DOI: 10.21248/hpsg.2023.7.

Abstract

Semantic Representations become useful resources for various multilingual NLP applications such as Machine Translation, Multilingual Generation, cross Lingual QA, to name a few. No Semantic Representation, to our knowledge, adopts *vivakṣā* (*Speaker's intention)* as a guiding principle for the representation. This motivates us to develop a new Semantic Representation system – **Universal Semantic Representation (USR)** – following Indian Grammatical Tradition (IGT) and Paninian grammar. Since USR is designed to be language-independent, we have currently taken up the task of generating English, Hindi, Tamil and Bangla from the USR. For English generation, the USR is mapped to ERG meaning representation (Flickinger, D. 1999) which is couched in Minimal Recursion Semantics (MRS). We use an off-the-shelf ACE generator that uses ERG as a resource-grammar for generating English. While designing the transfer module from USR to ERG-based MRS, we came across various Abstract Predicates (APs) in MRS representation as described in ErgSemantics_Basic (Flickinger et al., 2014). These APs are used to represent the semantic contribution of grammatical constructions or more specialized lexical entries such as compounding or the comparative use of more and so on. This paper presents the strategy for postulating the APs from the information given in USR and then reports the implementation of the transfer module keeping the focus on the postulation of APs. We get around 95% accuracy in postulating APs from USRs.

1. Introduction

One major advantage of Semantic Representations (SemRep) is the potential cross-linguistic *universality* (Abend and Rapparport, 2017) that these SemReps can ideally represent. Languages differ in terms of their form but they have often been assumed to be much closer in terms of their semantic content (Bar-Hillel, 1960; Fodor, 1975) and SemRep can capture that content. Thus Semantic Representations become useful resources for various multilingual NLP applications such as Machine Translation (MT) (Hajič 2002), Multilingual Generation (Cabezudo et al., 2019), cross Lingual QA, to name a few.

Generally, all SemReps abstract away from grammatical and syntactic idiosyncrasies inherent in natural languages (Boguslavsky et al., 2021). As is evident in Semantic Role Labeling (Gildea and Jurafsky, 2002), FrameNet (Baker et al., 1998), Propbank (Kingsbury and Palmer, 2002), Abstract Meaning Representation (Banarescu et al., 2013), the fundamental component of the content conveyed by SemReps of texts is argument structure – who did what to whom, where, when and why, i.e., events, their participants and the relations between them (Abend and Rapparport, 2017). However, in communication, speakers express through an utterance how (s)he views the situation which the mere argument structure of events can never capture. Thus what is expressed in communication is vivakṣā: the intention of the speaker about the meaning to be conveyed by the words. IGT views discourse composition as the manifestation of the speaker's vivakṣā. Example (1.a) and (1.b) explain how vivakṣā determines the syntactic expressions:

(1) a. **umā ko** kala rāta cāṃda dikhā
umā k4a yesterday-r6 night-k7t moon-k1 see(intr)-past
'umā happened to see the moon yesterday night'

b. ***umā ne*** *kala rāta cāṃda dekhā*
umā k1 yesterday-r6 night-k7t moon-k2 see(tr)-past
'*umā* saw the moon yesterday night'

The activity of 'seeing' licenses an animate *seer* and a *seen entity*. That is the *semantic frame* for the verb. However, in communication, it is not the semantic frame of a chosen event alone that depicts the situation. Two other important factors also play a major role: (a) how the speaker conceptually cognizes the situation? (b) which linguistic expressions are available to translate that cognition into languages? For example, in (1), does the speaker want to express Uma's agency or does (s)he want to foreground the moon's appearance over the seer's agency? This is termed the speaker's *vivakṣā*. Depending on that, the speaker would choose the most appropriate linguistic expressions to convey his/her thoughts. For example, the speaker views the situation very differently when (s)he says (1.a) '*Uma happened to see the moon yesterday night'* vis-à-vis (1.b) *'Uma saw the moon yesterday night'*. In Hindi, two different verb roots are used and the post-position on the seer also indicates different kāraka relations. In (1.a), Uma is an experiencer, while in (1.b), the volitionality of Uma is maintained.

To our knowledge, no SemRep adopts *Speaker's intention* as a guiding principle for the representation. This motivates us to develop a new Semantic Representation system – **USR** – following IGT (Sukhada et al., 2023) and Paninian grammar (Zdeněk Žabokrtský et al., 2020). The application task chosen is Multilingual Natural Language Generation. Since USR is designed to be language-independent, we have currently taken up the task of generating English, Hindi, Tamil and Bangla from the USR.

For English generation, the USR is mapped to ERG meaning representation (Flickinger, D. 1999) which is couched in MRS (Copestake et al., 2005). We use an off-the-shelf ACE generator that uses ERG as a resource-grammar for generating English. Since both USR and MRS are semantics-based representations, we assume that the USR-MRS transfer would be straightforward. While designing the transfer module from USR to ERG-based MRS, we came across various APs in MRS representation as described in ErgSemantics_Basic (Flickinger et al., 2014). These APs are used to represent the semantic contribution of grammatical constructions or more specialized lexical entries such as compounding or the comparative use of more and so on.

This paper presents the strategy for postulating the APs from the information given in USR and then reports the implementation of the transfer module keeping the focus on the postulation of APs.

Section 2 introduces the new SemRep USR briefly. Section 3 presents the similarities and differences between USR and MRS to motivate the significance of writing a transfer grammar. Section 4 describes the APs as postulated in ERG meaning representation. Section 5 discusses in detail the implementation of Transfer Grammar for APs in CLIPS. The experiment, results and error analysis for the task *generation of APs* is reported in Section 6. Finally Section 7 concludes the paper.

2. A Brief Introduction to USR

USR attempts to design a structured representation for the speaker's vivakṣā. IGT views language as a holistic phenomenon (Sukhada et al., 2023). Words are not derived as isolated units in Paninian grammar, but as units that are semantically connected with other words in the sentence (Raster, 2015). Sentences are connected across the discourse. This is explicitly recognized by the Paninian rule (A 2.1.1): *samarthaḥ padavidhiḥ*. Keeping in mind Natural Language Generation as the targeted application, the lexico-semantic and relational information is specified in the USR at various layers so that proper word forms, relations among words and finally relations across sentences can be generated systematically.

USR is a csv-formatted multilayered information packaging system that encapsulates (a) lexico-conceptual, (b) syntactico-semantic relational and (c) discourse level information (Garg et al., 2023). The uniqueness of this representation is that information on each layer is distinctly yet interactively maintained through attribute value matrix and co-referencing as shown in sentence (2). The USR for the semantics of sentence (2) is given in Table 1:

(2) hari ne apane guru jī ko garama dūdha aura miṭhāī dekara
hari erg his teacher respect dat hot milk and sweet offering
ābhāra vyakta kiyā
gratitude express do.pst
'hari expressed gratitude to his respected teacher by offering hot milk and sweet'

Astract layer	Con(cept)	hari	apanā /self	guru_1/ teacher_5	garama_1/hot_1	dūdha_1/milk_1	miṭhāī_1/sweet_1	de_17/ offer_1	ābhāra_1/ gratitude_1	vyakta+kara_1/ express_1-yA_1/past_1
Lexico-Conceptual	Sem(antic)_Cat(egory)	person male		anim						
	Morpho-Sem(antic)						plural			
	Speaker-view			respect						
	Index	1	2	3	4	5	6	7	8	9
Syntactico-Semantic	Dep(endency)	9:k1 (doer/agent)	3:r6 (genitive)	7:k4 (recipient)	5:mod	7:k2 (object)	7:k2	9:rpk (sequence of events)	9:k2 (theme)	0:main
	Construction	conjunction:[5,6]								
Discourse	Coreference & Connective		1:coref							
	Sent Type	affirmative								

Table 1: USR for sentence (2)

In Table 1, the **Concept row** represents *unique* concepts (not words) that refer to entities, events, quality, quantity and other properties of an entity or event. For sentence (2), the concepts are *Hari, apanā/self, guru_1/teacher_5,* *garama_1/hot_1, dūdha_1/milk_1, de_9/offer_1, miṭhāī_1/sweet_1, ābhāra_1 /gratitude_1, vyakta+kara_1/express_1*. TAM (Tense-aspect-modality) is also treated as a concept and therefore the TAM 'past' has been specified on the main verb. We have not considered *aura(Conjunction and), jī(respect marker)* as concepts in the concept row. The speaker intends to show *respect* to *his* teacher. Therefore the information *'respect'* has been specified in the **Speaker's view row** under the concept *guru_1/teacher_5*. During generation, the corresponding word *jī* in Hindi, *bābu* in Bangla, *gāru* in Telugu and the modifier 'respected' in English will be generated. Currently, the Semantic category row contains named entity information (C S & Lalitha Devi, WILDRE 2020), animacy and inherent gender information for the concepts. The Morpho-semantic row captures semantic information such as number, comparison and causation which can be marked in languages morphologically. Relations among concepts are specified in terms of dependency relation in the Dependency row. The Construction row conveys non-dependency relational information. For example, in a conjoined construction, all entities involved enjoy equal status. The discourse level information such as inter-sentential connectivity, co-referencing are represented in the Discourse row. Finally, the Sentence type is also specified.

3. Motivation for Transfer Grammar Module

As stated earlier, the main application task planned for USRs is multilingual generation. Since for English generation, the open-source ACE generator is available and the input it takes is a kind of semantic representation in MRS format, we examined if a transfer grammar module can be developed for converting USR to MRS. The advantage is that we would not be required to develop an English generator from scratch. Moreover, apart from English, already large-scale MRS-based grammar is available for a few other languages such as German, Japanese, and Korean. Thus in the future, those languages can also be generated from USRs via MRS. This section describes the similarities and differences found during examining the USR and MRS representations.

3.1 Similarities between USR and MRS representation

The motivating factor for writing a transfer module from USR to MRS is that there are many similarities between the two representations such as the following:

- Both USR and MRS are semantic representations that abstract way the syntactic idiosyncrasies of languages
- The finite verb is the head/root of the representation
- TAM (tense-aspect-modality) is represented as features on the verb
- GNP (gender-number-person) information are attested on the nouns.

- Adjectives are treated as a stand-alone concept even where they are derived from nouns or verbs.
- Verb-argument structure is specified in the representation
- No canonical representation. For example, active voice and passive voice sentences are represented as different semantic representations.

Thus principally, the conversion between USR and MRS can be direct. However, there are differences observed between the two representations, especially in terms of the postulation of APs in MRS, that necessitate a constraint-based transfer grammar module. The abstract predicate mapping that is the focus of the paper highlights the dissimilarity between the two frameworks as shown in Section 4. However this paper discusses differences related to APs alone.

4. Abstract Predicates in MRS

The predicate symbols in ErgSemantics have been divided into two classes: *surface* predicates and *AP*s. In non-lexical contexts, APs come into play, whether to represent ordinals such as "*first*" with "/*ord/*" or to denote negative constructions using "/*neg/*." ERG has around 108 APs. They can be classified into the following broad categories:

i. Quantifier
ii. Abstraction
    a. Degrees of Comparison
    b. Pronoun
    c. Named Entity
    d. Time and Place
    e. Question
    f. Number
iii. Construction
iv. Other

Table 2 shows the list of the APs handled so far:

<table>
<tr><th colspan="6">Quantifier</th></tr>
<tr><td colspan="2">/def_explicit_q/</td><td>/udef_q/</td><td colspan="2">/proper_q/</td><td rowspan="2">/every_q/</td></tr>
<tr><td colspan="2">/def_implicit_q/</td><td>/which_q/</td><td colspan="2">/pronoun_q/</td></tr>
<tr><th colspan="6">Abstraction</th></tr>
<tr><th>Degrees of Comparison</th><th>Pronoun</th><th>Named Entity</th><th>Time & Place</th><th>Question</th><th>Number</th></tr>
<tr><td>/comp/</td><td>/pron/</td><td>/named/</td><td>/loc_nonsp/</td><td>/measure/</td><td>/card/</td></tr>
<tr><td>/comp_equal/</td><td></td><td>/dofw/</td><td>/place_n/</td><td>/thing/</td><td>/ord/</td></tr>
<tr><td>/comp_less/</td><td></td><td>/mofy/</td><td>/time_n/</td><td>/reason/</td><td></td></tr>
<tr><td>/superl/</td><td></td><td>/yofc/</td><td></td><td>/property/</td><td></td></tr>
</table>

/abstr_deg/		/season/		/manner/	
				/person/	
Construction					
Compounds	**Passive**	**Non-Finite**	**Negation**	**Possession**	**Reciprocal Pronoun**
/compound/	/parg_d/	/subbord/ /nominalization/	/neg/	/poss/	/recip_pro/
Others					
/unspec_manner/					

Table 2: List of the APs handled so far

This paper attempts to identify where and how information encoded in USR enables to postulate the aforementioned APs. In most of the cases, semantic information encoded in USR is used to determine the APs while there are few cases where we are currently using mainly entries of Concept row to postulate APs. Table 3 to Table 6 specify which information from USR is being used to predict the right AP of different categories.

i. Quantification

As described in the ErgSemantics_Basic document, the ERG assumes that all instance variables (of type *x*) are bound by a generalized quantifier. Such an assumption is not taken in USR. Table 3 indicates the information that we are using from USR to postulate *_q APs. In column 3 of Table 3 to Table 6, the convention ('<' is used for *binds*, '|' for *when*, ':' *in* * row) is used.

MRS quantifier	**Context**	**Rules from USR information**	**Example**
/def_explicit_q/ (1)	Possessive nouns & pronouns	1 < noun$_i$ | *r6* for **i** : Dep row	Ram**'s** book
/def_implicit_q/ (2)	Spatial & temporal adverbs	2 < here & /place/ | *yahāṃ* : Con row 2 < there & /place/ | *vahāṃ* : Con row 2 < now & /time/ | *aba* : Con row 2 < today & /time/ | *āja* : Con row 2 < tomorrow & /time/ | *kala* : Con row *2* < /poss/, /person/ & /which_q/ | *kim*$_i$ *:* Con row + *r6* for **i** : Dep row	The boy lives **here.** I am going **there.** He will come **now.** The meeting is **today.** She will catch the train **tomorrow.** **Whose**$_i$ house is this?
/every_q/ (3)	Universal	3 < /person/ | *saba* : Con	Rama calls **eve-**

	quantifier	row	**rybody** in the school..
/proper_q/ (4)	Proper noun	4 < /named/ | *per :* Sem_Cat row 4 < /named/ | *place :* Sem_Cat row 4 < /dofw/ | *dow :* Sem_Cat row 4 < /mofy/ | *mofy:* Sem_Cat row *4* < /yofc/ | *yofc :* Sem_Cat row	**Sanju** is good. **India** is a sub-continent. Babies eat fruits on **Monday.** **January** is the first month. He will come to India in **2024.**
/pronoun_q/ (5)	Personal pronoun	5 < /pron/ | *speaker :* Con row + *+sg/pl :* Morpho-Sem_row 5 < /pron/ | *addressee :* Con row *+sg/pl :* Morpho-Sem_row 5 < /pron/ | *3rd person/wyax:* Con row + *coref :* discourse_row	**We** are going to a party. **You** are a good person. **He** is smart.
/which_q/ (6)	Interrogative pronoun	6 < /person/ | kim_i : Con row + *k1* for **i** : Dep row + anim : Sem_Cat row 6 < /time/ + /loc_nonsp/ | kim_i : Con row + *k7t* for **i** : Dep row 6 < /place/ + /loc_nonsp/ | kim_i : Con row +*k7p*/*k2p* for **i** : Dep row 6 < /thing/ | kim_i : Con row + *k2* for **i** : Dep row - animacy : Sem_Cat row 6 < /reason/ | kim_i : Con row + rh for **i** : Dep row 6 < /manner/ & /unspec_manner/ | kim_i : Con row + *krvn*: Dep row	**Who**$_i$ filled the bottle? **When**$_i$ will you come? **Where**$_i$ are you going? **What**$_i$ are you buying? **Why**$_i$ are you sad? **How** did you finish the work?
/abstr_deg/ (7)	Interrogative Degree	7 < /measure/ & /which_q/ | kim_i : Con row + *degree* relation for **i** : Dep row + *interrogative* : Sent_Type row	**How**$_i$ **happy** was Sita?

Table 3: MRS quantifiers from the USR information utilized

ii. Abstraction

This category consists of cases where MRS representation goes one level more abstract than the surface predicates to capture certain generalization in the representation, example /*comper_equal/* for the similarity between two different entities. Similarly APs for the named entities, adverbs of time, and numerals as well as the information specified in the USR for these predicates have also been listed in Table 4:

MRS Predicates	**Context**	**Rules from USR information**	**Example**
/comp/ (8)	Comparative degree more	ARG1 of 8 is adj$_i$ & ARG2 is noun$_j$ | *compermore* of **i** : Mor_Sem_row + *rv* for **j** : *Dep row*	Sanju is **more** intelligent$_i$ than Rahul$_j$.
/comp_less/ (9)	Comparative degree less	ARG1 of 9 is adj$_i$ & ARG2 is noun$_j$ | *comperless* of **i** : Mor_Sem_row + *rv* for **j** : *Dep row*	Mohan is **less** intelligent$_i$ than Rama$_j$.
/comp_equal/ (10)	Similarity	ARG1 of 10 is adj$_i$ & ARG2 is noun$_j$ | *ru* relation for **j** : Dep row	Sita is **as beautiful**$_i$ **as** Radha$_j$.
/superl/ (11)	Superlative degree	ARG1 of 11 is adj$_i$ | *superl* of **i** : Morpho_Sem_row + *k1s* for **j** : Dep row	The sun is the **biggest**$_i$ star$_j$.
/pron/ (12)	Personal pronouns	12 | *speaker:* Con row + *sg/pl:* Mor-Sem_row 12 | *addressee:* Con row +*sg/pl:* Mor-Sem_row 12 | *wyax:* Con row + *coref :* discourse_row	**I** bought a diary. **You** are smart. **They** will go to Banaras.
/named/ (13)	Proper noun	13 | *per :* Sem_Cat row 13 | *place :* Sem_Cat row	**Rama** ate an apple. **Rama** lives in Ayodhya.
/dofw/ (14)	Name of the days of week	14 | *dow :* Sem_Cat row	**Sunday** is a holiday.
/mofy/ (15)	Name of the months of year	15 | *mofy:* Sem_Cat row	**December** is the final month of the year.

/yofc/ (16)	Year of centuries	16 | *yofc:* Sem_Cat row	What will happen in **2025**?
/season/ (17)	Name of the seasons	17 | *season* : Con row	Christmas is celebrated in **winter**.
/loc_nonsp/ (18)	Spatial & temporal entities	18 | *yahāṃ* : Con row 18 | *vahāṃ* : Con row 18 | *aba* : Con row 18 | *āja* : Con row 18 | *kala* : Con row	He lives **here.** I will be **there** in five minutes. The teacher will teach **now.** He is happy **today.** **Tomorrow** is a holiday.
/place_n/ (19)	Spatial entities	19 | *kim*$_i$: Con row + *k7p*/*k2p* for i : Dep row 19 | *yahāṃ* : Con row 19 | *vahāṃ* : Con row 19 | *kim*$_i$: Con row + *k5* for i : Dep row	**Where**$_i$ do you live? Kids are **here**. Your bicycle is **there.** **Where**$_i$ did you come from?
/time_n/ (20)	temporal adverbs	20 | *aba* : Con row 20 | *āja* : Con row 20 | *kala* : Con row 20 | *kim* : Con row + *k7t* for **i** : Dep row	She is reading the book **now.** He plays the guitar **today.** We will buy groceries **tomorrow.** **When**$_i$ are you leaving?
/measure/ (21)	Abstract Measuring	ARG1 of 21 is adj$_i$ & ARG2 is which_q | *kim*$_j$: Con row + *degree* relation for **j** : Dep row + *interrogative* : Sent_Type row	**How**$_j$ sad$_i$ was Sita?
/thing/ (22)	Wh - word "*What*"	22 | *kim*$_i$: Con row+ *k2* of **i** : Dep row	**What** are you doing?

<table>
<tr><td>/property/ (23)</td><td>How are you?</td><td>23 | kim_i : Con row + k1s for i : Dep row + interrogative : Sent_Type row + animacy : Sem_Cat row</td><td>How$_i$ are you?</td></tr>
<tr><td>/person/ (24)</td><td>wh_words with animacy</td><td>24 | kim_i : Con row + k1 for i : Dep row + animacy : Sem_Cat row
24 | kim_i : Con row + k1s for i : Dep row + animacy : Sem_Cat row
24 | kim_i : Con row + k2/k2g/k4 for i : Dep row + animacy : Sem_Cat row
24 | kim_i : Con row + r6 for i : Dep row
24 | kim_i : Con row + k5 for i : Dep row + animacy : Sem_Cat row</td><td>Who$_i$ finished the work?

Who$_i$ is Rama?

Whom$_i$ did Rama meet?

Whose$_i$ car is that?

Who$_i$ is Mohan afraid of?</td></tr>
<tr><td>/reason/ (25)</td><td>Why word</td><td>25 | kim_i : Con row + rh for i : Dep row + interrogative : Sent_Type row</td><td>Why$_i$ are you crying?</td></tr>
<tr><td>/manner/ (26)</td><td>Interrogative Manner</td><td>26 | kim_i : Con row + krvn for i : Dep row + interrogative : Sent_Type row</td><td>How$_i$ did you come?</td></tr>
</table>

Table 4**:** Generic APs with Examples

iii. Construction

This category includes what we commonly call as construction, form-meaning pairs. For different constructions different kinds of information from USR is being utilized for the mapping.

<table>
<tr><th>MRS predicates</th><th>Context</th><th>Rules from USR information</th><th>Example</th></tr>
<tr><td>/compound/ (27)</td><td>Compound words & English Honor-</td><td>27 | $noun_j$+$noun_i$: Con row</td><td>He laid the foundation$_j$</td></tr>
</table>

	ific words	27 | noun$_i$: Con row; *respect* : Speaker_View row of noun$_i$	**stone**$_i$ for Rama's office. **Ms**$_j$**. Sita**$_i$ joined the course.
/parg_d/ (28)	(i) Passive sentences & (ii) rbks	(i) 28 | *Passive TAM* : Con row + *passive* : Sent_Type row (ii) 28 | verb$_i$: Con row + rbks for **i** : Dep row + *affirmative* : Sent_Type row	Ravana$_j$ was **killed**$_i$ **by** Rama. The fruit **eaten**$_i$ **by** Rama$_j$ was sweet.
/subord/ (29)	Subordinate clauses	(i) 29 | verb$_i$: Con row + *rpk* for **i**: Dep row (ii) 29 | verb$_i$: Con row + *krvn* for **i** : Dep row	Having been **going**$_i$ to the school, Rama **ate** food. Mohan walks **limping**$_i$ **.**
/nominalization/ (30)	Nominalized	30 | verb$_i$: Con row + k1 of : Dep row	**Chasing**$_i$ the cat is old.
/neg/ (31)	Negation	31 | *neg* : Dep row	I am **not** going$_i$ to the function.
/poss/ (32)	possession	32 | *kim*$_i$: Con row + *r6* of i : Dep row 32 | *noun*$_i$: Con row + *r6* of i : Dep row	**Whose** idea is this? I borrowed **Rama's cycle**.
/recip_pro/ (33)	Reciprocal Pronouns	33 | *eka* + *dūsarā*$_i$: Con row	Rama and Sita like each other$_i$**.**

Table 5: APs under construction category

iv. Other

This category includes the APs which are unique for interrogative pronouns. For different types of interrogative pronouns different kinds of information from the USRs is utilized for the mapping.For instance, when the USR has *kim* in the Concept row, *krvn* relation in Dep row and *interrogative* in

Sent_Type then the AP /*unspec_manner/* will be postulated in the transfer module as shown in Table 6.

MRS quantifier	Context	Rules	Example
unspec_manner (34)	Interrogative Manner	34 | *kim*$_i$: Con row + *krvn* of **i** : Dep row + *interrogative* : Sent_type	**How**$_i$ **did you complete your work?**

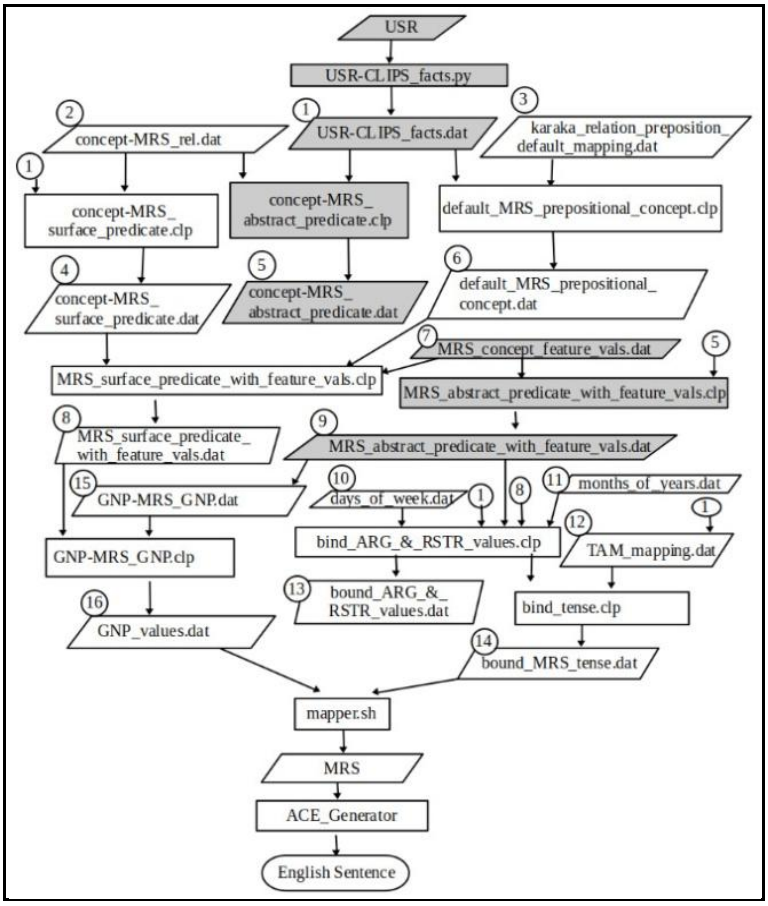
Table 6**:** Other APs

5. Implementation of Transfer Grammar for Abstract Predicates in CLIPS

The implementation is done at two levels: (a) Determining an AP (b) Specifying the feature structure description of the AP. For (a), information from USR has been utilized as discussed in the previous section (see Table 3 - Table 6). Once APs are identified for a given USR, we translate the sentence into English and run the ACE parser to find out the feature structure description of the targeted AP and add the AP along with its feature structure description in the dictionary if it is not already present there. This is our development stage for populating APs into the dictionary with appropriate feature structure description. Thus the lexicon for APs is created.

During English sentence generation from USR via MRS, this dictionary is consulted for framing the appropriate MRS for a given USR which in turn is used by ACE generator as an input and the English sentence is generated. The postulation of APs from USR is executed in CLIPS (Giarratano, J. C. 1993). The part with gray background of the flow chart describes the postulation of APs.

The USR_to_CLIPS_facts.py program converts a USR into CLIPS facts: CLIPS_ facts.dat (1). The concept-MRS_abstract_predicates.clp implements the rules (Table 3 - Table 6) and postulates AP types in concept-MRS_abstract_predicate.dat (5). Finally, the program MRS_abstract_predicate_with_feature_vals.clp takes two files as input, one dictionary MRS_concept_ feature_vals.dat (7) and the output file (5), and returns MRS_abstract_ predicate_with_feature_vals.dat(9) that contains all APs with their feature structure description.



Flowchart 1: USR to MRS transfer module (APs in gray background)

We will explain Flowchart 1 with an example (3):

(3) īśā ji ne apane beṭe aura apanī beṭī ko somavāra ko
Isha respect k1 her son-k2 and her daughter k2 Monday k7t
kāśī ke sabase baḍe vidyālaya meṃ bharatī kiyā.
Kashi-place r6 most large school k7p admit do-past
'Ms. Isha admitted her son and her daughter, on Monday, in Kashi's largest school.'

#īśā ji ne apane beṭe aura apanī beṭī ko somavāra ko kāśī ke sabase baḍe vidyālaya meṃ bharatī kiyā.										
Con-cept	īśā	apanā/her	beṭā_1/son_1	apanā/her	beṭī_1/daughter_1	somavāra/Monday	kāśī	baḍā_2/big_1	vidyālaya_1/school_1	bharatī + kara_6-yā_1/admit_1-past

Index	1	2	3	4	5	6	7	8	9	10
Sem_Cat	per fe ma le		anim male		anim female	dow	pl ac e			
Morph_Sem								superl		
Dep	10: k1	3:r6	10:k2	5:r6	10:k2	10:k7t	9: r6	9:mod	10:k7p	0:main
Dis-course		1:co ref		1:co ref						
Speak er-view	re- spe ct									
Con-structi on	conjunction:[3,5]									
Sent_Type	affirmative									

Table 7: USR for the sentence (3)

The rules for generating MRS APs from the information specified in USR are listed in Table 8 for the concepts given in the first column. For example, two APs occur for any named entity in MRS. For īśā, USR specifies per: yes (i.e., īśā is a person) which generates the AP /named/ which is bound by the quantifier /proper_q/. The ‘yes’ value for per (in Sem_Cat row), female (in Morpho_Sem row) and respect (in speaker’s view row generate two APs /compound/ and the quantifier /udef_q/ which in turn generates the lexical item “Ms.” as a compound “Ms. Isha”. Similarly we can examine the USR information utilized for generating APs in the context of other concepts in Table 8.

Concept label	**Eng. Eqiv.**	**Information specified in USR**	**Generating AP/s**	**Purpose**
īśā(1)	*īśā*	*per:yes*	/named/, /proper_q/,	*īśā*
		per:yes, respect:yes, feminine:yes	/compound/, /udef_q/,	Ms.
apanā(2)	*she*	*coreference to* 1, *feminine:yes*	/pronoun_q/,/pron/	her
		genitive to 3	/poss/, /def_explicit_q/	

		conj: [3,5]	/udef_q/	son and daughter
beṭā_1(3)	*son*			
apanā(4)	*she*	*coreference* to 1, *feminine:yes*	/pronoun_q/, /pron/	her
		genitive to 5	/poss/, /def_explicit_q/	
beṭī_1(5)	*daughter*			
somavāra(6)	*Monday*	*dow:yes*	/dofw/, /proper_q/	Monday
kāśī(7)	*kāśī*	*place:yes*	/proper_q/, /named/	*kāśī*
		genitive relation to noun (here *school*)	/def_explicit_q/, /poss/	*kāśī*'s
baḍā_2(8)	large	superl to 8:yes	/superl/	largest

Table 8**:** FlowChart 1 explained using example (3)

Surface predicates are handled separately using another CLIPS program. We assign feature values to surface predicates from the MRS_concept_feature_vals.dat(7) dictionary. After incorporating GNP values from the GNP dictionary, ARGument sharing will be done. Following this, binding of handle constraints LBL and RSTR values is done. Subsequently, mapper.sh will run for obtaining the complete MRS representation of a USR. The MRS representation then becomes input to the ACE generator for producing natural English sentences.

5.1 Statistical Observation on Transfer Rules

The implementation of rules for creating the APs include three types of mapping:

1. Direct Mapping: A relation or a lexical concept from USR is directly mapped to MRS AP;
2. Indirect Mapping: Information encoded at multiple layers in USR is used to postulate the AP;
3. Constraint based mapping: Where the rule includes constraints to prevent wrong or overgeneration of APs.

Examples for each type of mapping are given in Table 9 to Table 11.

USR information utilized	**Context**	**Example**	**MRS AP**
Morpho-semantic row	Superlative degree	*superl*	/superl/
Dependency relation row	Sequence of events	*rpk*	/subord/

Table 9: Examples of Direct Mapping

USR info utilized	**Context**	**Ex.**	**MRS AP**
Concept & de-	Interrogative pro-	*where*	/which_q/, /loc_nonsp/, /place_n/

pendency	noun		
Concept & dependency	Implicit quantifier	*whose*	/def_implicit_q/, /poss/, /person/, /which_q/

Table 10: Examples of Indirect Mapping

In Constraint-based mapping, we take into account the mapping rules that apply some constraints for generating an AP. For example, the generation of /person/ and /thing/ use the same information, *kim* in Con row and *k2* relation in Dep row. The distinguishing factor that works is the *animacy* feature in Sem_Cat row. The presence of an *animacy* feature triggers the postulation of /person/, that along with /which_q/ generates '*who'* in English sentences. On the other hand, the absence of an *animacy* feature in Sem_Cat row postulates /thing/ which along with /which_q/ generate the English word '*What'.*

USR information utilized	**Context**	**Example**	**MRS AP**
Concept row	*kim*	*vahāṁ*	/place_n/, /def_implicit_q/, /loc_nonsp/
Semantic Category row	*per/place/org/ne*	*Sanju*	/proper_q/, /named/

Table 11: Examples of Constraint-based Mapping

Rule	**Number of Rules**	**Percentage**
Direct mapping	10	27.78%
Indirect mapping	15	41.67%
Constraint based mapping	11	30.55%

Table 12: Statistical observation on transfer rules in CLIPS

We observe that rules written for Indirect mapping are the highest in number. Thus, we conclude that information used for postulating MRS APs is distributed at different layers of the USRs.

6. Experimental Setup, Result and Error Analysis

Preparing a test suite for APs is challenging. Although experienced linguists have been given the task, they do not have an idea of APs which is a framework internal feature of MRS. To address this issue, we have prepared a short guideline for the linguists who created the test suite. For each AP, we created 5 sample sentences with the word under consideration underlined and asked the linguist to create 10 more sentences in which the underlined words must be present. For example, for /card/ one of the 5 sample sentences was 'the boy ate five mangoes'. For some cases where instruction statements can easily be prepared, we have given the instruction statement along with 5 sample sentences. For example, for person names, we have given one sample sentence, 'Hari came home', and also the following instruction: "The sentence should have a person name”. Following the guidelines, our linguist team has created 262 test suite sentences. USR annotators were then asked to create USRs for these sentences. After obtaining the USRs, we proceeded to execute them using the transfer module and ACE Generator. Following this, we meticulously examined whether the anticipated APs were accurately generated. The Results and Error analysis has been given in Table 13.

Total USRs	**Total expected APs**	**APs generated**	**Error Analysis**	
			Concept missing	**Typographical error**
262	491	469	16	6
		95.5%	3.25%	1.22%

Table 13: Result and Error Analysis for APs

The result shows quite a promising conversion rate from USR to MRS as far as APs are concerned. The errors occur mainly due to wrong USR input as indicated in Table 13.

Table 14 shows results for each class of APs.

	Quantifiers	**Abstraction**	**Construction**	**Others**	**Total APs**
	110	281	70	30	**491**
Total errors	10	8	2	2	**22**
Accuracy	90.9%	97.15%	97.14%	93.33%	**95.5%**

Table 14 Results for each class of APs

7. Conclusion

This paper presents an architecture and implementation of converting the semantic representation USR to another semantic representation MRS to generate Natural language English using the open-source ACE generator. The focus of the paper has been on postulating APs which is a theory internal construct of MRS. USR is based on Indian Grammatical Tradition and Panini, while MRS is rooted in HPSG. It is interesting to note that USR does capture almost complete information that APs tend to represent. That is why we get 95% accuracy in postulating APs from USR. The only exception is the Quantifier APs of MRS. USR does not work with the assumption that every noun or noun phrase will have to be bound by a quantifier. Nevertheless, we were successful in generating all *_q predicates. The result of the work is surely motivating enough to develop a full-fledged transfer grammar module from USR to MRS for English and other languages as well for which MRS-based grammar exists.

References

Abend, O., & Rappoport, A. (2017, July). The state of the art in semantic representation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 77-89).

Bharati, A., & Kulkarni, A. (2008). Information Coding in a language: Some insights from Paninian Grammar. Dhimahi, Journal of Chinmaya International Foundation Shodha Sansthan I (1), 77–91

Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. International Journal of Lexicography, 18(4), 445-478.

Boguslavsky, I. Semantic analysis supported by inference in a functional model of language [Semanticheskij analiz s oporoj na umozakljuchenija v funktsional’noj modeli jazyka]. Problems of linguistics [Voprosy jazykoznanija], (1-2021), 29-56.

Cabezudo, M. A. S., & Pardo, T. A. S. (2019). Natural language generation: recently learned lessons, directions for semantic representation-based approaches, and the case of Brazilian Portuguese language. In Proceedings.

Copestake, A., Flickinger, D., Malouf, R., Riehemann, S., & Sag, I. (1995). Translation using minimal recursion semantics. In Proceedings of the Sixth Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages.

Copestake, A., Flickinger, D., Pollard, C., & Sag, I. A. (2005). Minimal recursion semantics: An introduction. Research on language and computation, 3, 281-332.

Flickinger, D., Bender, E. M., & Oepen, S. (2014, May). Towards an Encyclopedia of Compositional Semantics: Documenting the Interface of the English Resource Grammar. In LREC (pp. 875-881).

Garg, K., Paul, S., Sukhada, Kumari, R., and Bahawir, F. 2023. Evaluation of Universal Semantic Representation (USR). Proceedings of the 4th International Workshop on Designing Meaning Representations. 2023.

Giarratano, J. C. (1993). CLIPS user's guide. NASA Technical Report, Lyndon B Johnson Center.

Hajic, J. (2002, May). Tectogrammatical representation: towards a minimal transfer in machine translation. In Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+ 6) (pp. 216-226).

Kim, J. B. (2006). Minimal recursion semantics: An application into Korean. Journal of the Linguistic Association of Korea, 14(2), 59-85.

Kiparsky, P. (1995). Paninian Linguistics. In Concise History of the Language Sciences (pp. 59-65). Pergamon.

Kulkarni, A. (2008). Panini's Ashtadhyayi A Computer Scientist's viewpoint, 4th APCAP NIAS Bangalore

Malarkodi C S and Sobha Lalitha Devi. 2020.A Deeper Study on Features for Named Entity Recognition. In Proceedings of the WILDRE5– 5th Workshop on Indian Language Data: Resources and Evaluation, pages 66–72, Marseille, France. European Language Resources Association (ELRA).

Ranta, A., Angelov, K., Gruzitis, N., & Kolachina, P. (2020). Abstract syntax as interlingua: Scaling up the grammatical framework from controlled languages to robust pipelines. Computational Linguistics, 46(2), 425-486.

Song, L., Zhang, Y., Wang, Z., & Gildea, D. (2018). A graph-to-sequence model for AMR-to-text generation. arXiv preprint arXiv:1805.02473.

Sukhada, Paul, S., and Bhattacharjee, B. 2023. Theoretical Foundation of Universal Semantic Representation. IIIT/TR/2023/11142.